BRAINWARE UNIVERSITY

Term End Examination 2018 – 19

Programme –Master of Technology in Computer Science and Engineering

Course Name – Data Science

Course Code – PEC-MCS201B

(Semester – 2)

Time allotted: 3 Hours   **Full Marks: 70**

[The figure in the margin indicates full marks. Candidates are required to give their answers in their own words as far as practicable.]

Group –A

(Multiple Choice Type Questions)   10 x 1 = 10

1. *Choose the correct alternative from the following*

(i)   Which of the following is performed by Data Scientist?

a.	Define the question	b.	Create reproducible code
c.	Challenge results	d.	All of the Mentioned

(ii)   Which of the following is required to implement a literate programming system?

a.	A programming language like Perl	b.	A programming language like Java
c.	A programming language like R	d.	All of the Mentioned

(iii)   Which of the following is most important language for Data Science?

a.	Java	b.	Ruby
c.	R	d.	Python

(iv)   Which of the following model is usually gold standard for data analysis?

a.	Inference	b.	Summarizing
c.	Subsetting	d.	None of above

(v)   Which of the following is a categorical outcome?

a.	Inferential	b.	Descriptive
c.	Causal	d.	All of above

(vi)   Which of the following approach should be used if you can't fix the variable ?

      a. randomize it   b. non stratify it

      c. generalize it   d. none of the Mentioned

(vii)   Which of the following is commonly referred to as 'data fishing' ?

      a. Data bagging   b. Data booting

      c. Data merging   d. None of the Mentioned

(viii)   Which of the following analysis are incredibly hard to infer?

      a. postProcess   b. preProcess

      c. Process   d. All of above

(ix)   Which of the following is one of the key data science skill ?

      a. Statistics   b. Machine Learning

      c. Data Visualization   d. All of the Mentioned

(x)   Which of the following is useful way to put text, code, data, output all in one document?

      a. Literate statistical programming   b. Object Oriented programming

      c. Descriptive programming   d. Query processing Language

Group – B

(Short Answer Type Questions)   3 x 5 = 15

Answer any *three* from the following

2.   What is Machine Learning? Classify Machine Learning Procedure.   [2+3]

3.   Describe about the various classification algorithm.   [5]

4.   What is the role of Activation Function? Write down the various activation functions.   [2+3]

5.   Differentiate Supervised Learning and Unsupervised Learning using suitable example.   [5]

6.   A certain couple tells you that they have two children, at least one of which is a girl. What is the probability that they have two girls?   [5]

Group – C

(Long Answer Type Questions)   3 x 15 = 45

Answer any *three* from the following

7.	(a)	What is Data Science? Define the role of a Data Scientist.	[2+3]
	(b)	How is Business Intelligence different from Data Science?	[5]
	(c)	What do you understand by the term Normal Distribution?	[5]
8.	(a)	How does data cleaning play a vital role in text analysis?	[4]
	(b)	Differentiate between univariate, bivariate and multivariate analysis.	[5]
	(c)	What is Cluster Sampling?	[3]
	(d)	What is Systematic Sampling?	[3]
9.	(a)	What are Eigenvectors and Eigenvalues?	[4]
	(b)	Can you cite some examples where a false positive is important than a false negative?	[6]
	(c)	Can you cite some examples where a false negative important than a false positive?	[5]
10.	(a)	What is logistic regression? State an example when you have used logistic regression recently.	[2+4]
	(b)	What are Recommender Systems?	[4]
	(c)	What are the various steps involved in an analytics project?	[5]
11.	(a)	Data science depends on a diverse set of skills. Shown the set of skill set required by the Data Science using Venn Diagram and explain in short.	[5]
	(b)	Describe how Back Propagation technique is used as data classifier.	[5]
	(c)	Explain Gradient Descent algorithm.	[5]